On the Distortion of the Eigenvalue Spectrum in MIMO Amplify-and-Forward Multi-Hop Channels

Jörg Wagner and Armin Wittneben

Communication Technology Laboratory, ETH Zurich

Sternwartstrasse 7, 8092 Zurich, Switzerland

{wagner, wittneben}@nari.ee.ethz.ch

Abstract

Consider a wireless MIMO multi-hop channel with n_s non-cooperating source antennas and n_d fully cooperating destination antennas, as well as L clusters containing k non-cooperating relay antennas each. The source signal traverses all L clusters of relay antennas, before it reaches the destination. When relay antennas within the same cluster scale their received signals by the same constant before the retransmission, the equivalent channel matrix $\mathbf{H}$ relating the input signals at the source antennas to the output signals at the destination antennas is proportional to the product of channel matrices $\mathbf{H}_l$, $l = 1, \ldots, L+1$, corresponding to the individual hops. We perform an asymptotic capacity analysis for this channel as follows: In a first instance we take the limits $n_\mathrm{s} \to \infty$, $n_\mathrm{d} \to \infty$ and $k \to \infty$, but keep both $n_\mathrm{s}/n_\mathrm{d}$ and k/n_d fixed. Then, we take the limits $L \to \infty$ and $k/n_\mathrm{d} \to \infty$. Requiring that the $\mathbf{H}_l$'s satisfy the conditions needed for the Marčenko-Pastur law, we prove that the capacity scales linearly in $\min\{n_\mathrm{s}, n_\mathrm{d}\}$, as long as the ratio k/n_d scales at least linearly in L. Moreover, we show that up to a noise penalty and a pre-log factor the capacity of a point-to-point MIMO channel is approached, when this scaling is slightly faster than linear. Conversely, almost all spatial degrees of freedom vanish for less than linear scaling.

I. INTRODUCTION

We consider coherent wireless multiple-input-multiple-output (MIMO) communication between n_s non-cooperating source antennas and n_d fully cooperating destination antennas. In this paper it is assumed that the source antennas are either far apart or shadowed from the destination antennas. The installation of intermediate nodes that relay the source signals to the destination (multi-hop) is well known for being an efficient means for improving the energy-efficiency of the communication system in this case. In the resulting network, the signals traverse L clusters containing k relay antennas each, before they reach the destination. Generally, signals transmitted by the source antennas might not only be received by the immediately succeeding cluster of relay antennas, but possibly also by clusters that are farther away or by the destination. While such receptions could well be exploited for achieving higher transmission rates, we assume them to be strongly attenuated and ignore them in this paper.

In the most basic MIMO multi-hop network architecture, the relay antennas in the clusters do not cooperate. Since non-cooperative decoding of the interfering source signals at the individual relay antennas drastically reduces the achievable rate in the network, a simple amplify-and-forward operation becomes the relaying strategy of choice. That is, at each antenna a scaling of the received signals by a constant is performed before the retransmission. While this approach is cheap in terms of computational complexity, and also does not require any channel-state information at the relay nodes, it clearly suffers from noise accumulation. This basic network has been studied extensively by Borade, Zheng and Gallager for independent identically distributed (i.i.d.) Rayleigh fading channel matrices. In references [1], [2] they showed for $n \triangleq n_\mathrm{s} = n_\mathrm{d} = k$ that all n spatial degrees of freedom are available in this network for a fixed L at high signal-to-noise ratio (SNR). More generally, they also showed that all degrees of freedom are available, if L as a function of the SNR fulfills $\lim_{\mathsf{snr}\to\infty} L(\mathsf{snr})/\log \mathsf{snr} = 0$.

While this result gives a design criterion how the SNR should be increased with the number of hops in the network, it does not give any insights into the eigenvalue distribution of the product of random matrices $\mathbf{C}$ specifying the mutual information between input and output of the vector channel. For fixed

L, only recently this eigenvalue distribution has been characterized in the large antenna limit [3]. Based on a theorem from large random matrix theory [4], the authors showed that it converges to a deterministic function, and gave a recursive formula for the corresponding Stieltjes transform. Moreover, the reference reports that the asymptotic eigenvalue distribution of $\mathbf{C}$, which is in fact the product of the signal covariance matrix $\mathbf{R}_s$ and inverse noise covariance matrix $\mathbf{R}_n^{-1}$ at the destination, approaches the Marčenko-Pastur law in the large dimensions limit for $\beta_r \triangleq k/n_{\mathrm{d}} \to \infty$, but $\beta_s \triangleq n_{\mathrm{s}}/n_{\mathrm{d}}$ and L fixed. Since the Marčenko-Pastur law is also the limiting eigenvalue distribution of the classical point-to-point MIMO channel, this means that up to a noise penalty and a pre-log factor the point-to-point capacity is approached in this case.

By considering the limiting eigenvalue distributions of the signal and noise covariance matrices separately, we are able to generalize this result for the case $L \to \infty$ in this paper. In essence, we show that β_r needs to grow at least linearly with L in order to sustain a non-zero fraction of the spatial degrees of freedom in the system, i.e., linear capacity scaling in $\min\{n_{\mathrm{s}}, n_{\mathrm{d}}\}$. Moreover, when the scaling is faster than linear, the limiting eigenvalue distribution of $\mathbf{C}$ is given by the Marčenko-Pastur law. That is, we are able exploit the spatial degrees of freedom without increasing the SNR at the receiver at the expense of employing more relay antennas. Returning to the result by Borade et al., where degrees of freedom are sustained by increasing the SNR, according to our result the number of relays per layer can be seen as a second resource besides the transmit power for compensating the capacity loss in the multi-hop network.

Another contribution of this paper lies in bridging the gap between the results obtained by Müller in reference [5] on the one hand and by Morgenshtern and Bölcskei in references [6], [7] on the other hand. In the first reference, it is shown in the large dimensions limit that almost all singular values of a product of independent random matrices fulfilling the conditions needed for the Marčenko-Pastur law go to zero as the number of multipliers grows large, while the aspect ratios of the matrices are kept finite. This implies that almost all spatial degrees of freedom in a MIMO amplify-and-forward multi-hop network as described above vanish as L goes to infinity. On the other hand, [6], [7] were the first papers which proved

in the large dimensions limit (for $L = 1$) that the capacity of a point-to-point MIMO link is approached up to a noise penalty and a pre-log factor, if $\beta_r \to \infty$ and β_s is kept fixed. In [8] the same result had been proven for the less general case that n_s and n_d are fixed and $k \to \infty$. The mechanisms discovered in these papers apparently act as antipodal forces with respect to the limiting eigenvalue distributions of products of random matrices. While increasing the number of hops distorts this distribution in an undesired fashion, increasing the ratio between the number of relays and destination antennas allows for recovering the original distribution corresponding to a point-to-point channel. In this paper, we answer the question how these two effects can be balanced, i.e., how fast must β_r grow with L in order to sustain a non-zero fraction of spatial degrees of freedom as L grows without bounds.

II. Notation

The superscripts ^{H} and * stand for conjugate transpose and complex conjugate, respectively. E_A denotes the expectation operator with respect to the random variable A. $\det(\mathbf{A})$, $\mathrm{Tr}(\mathbf{A})$ and $\lambda_i\{\mathbf{A}\}$ stand for determinant, trace and the ith eigenvalue of the matrix $\mathbf{A}$. $a(i)$ is the ith element of the vector $\mathbf{a}$. Throughout the paper all logarithms, unless specified otherwise, are to the base e. $\|\mathbf{a}\|$ denotes the Euclidean norm of the vector $\mathbf{a}$, $\|\mathbf{A}\|_{\mathrm{Tr}}$ the Trace norm of the matrix $\mathbf{A}$. By $\mathsf{Pr}[A]$ we denote the probability of the event A.

Furthermore, we use the standard $\mathcal{O}(\cdot), \Omega(\cdot), \Theta(\cdot)$ notations for characterizing the asymptotic behavior of some function $f(\cdot)$ according to

$$f(n) \in \mathcal{O}(g(n)) \text{ if } \exists M, n_0 > 0 : M|g(n)| > |f(n)|, \forall n \geq n_0,$$

$$f(n) \in \Omega(g(n)) \text{ if } \exists M, n_0 > 0 : M|g(n)| < |f(n)|, \forall n > n_0,$$

$$f(n) \in \Theta(g(n)) \text{ if } f(n) \in \mathcal{O}(g(n)) \text{ and } f(n) \in \Omega(g(n)).$$

Finally, we define the function $1\{x\}$ to be 1 if x is true and zero otherwise. $\delta(x)$ and $\sigma(x)$ denote Dirac delta and Heaviside step function, respectively.

III. TRANSMISSION PROTOCOL & SYSTEM MODEL

We label the clusters of relay antennas by $\mathcal{C}_1, \ldots, \mathcal{C}_L$. Cluster $\mathcal{C}_1$ denotes the one next to the destination, $\mathcal{C}_L$ the one next to the sources (refer to Fig. 1). We assume the $L+1$ single hop channels between sources, relay clusters and destinations to be frequency-flat fading over the bandwidth of interest, and divide the transmission into $L+1$ time slots. In time slot $T = 1$ the sources transmit to $\mathcal{C}_L$. The transmission is described by the transmit vector $\mathbf{s} \in \mathbb{C}^{n_\mathrm{s}}$, the matrix $\mathbf{H}_{L+1} \in \mathbb{C}^{k \times n_\mathrm{s}}$, representing the vector channel between sources and $\mathcal{C}_L$, the vector $\mathbf{n}_L \in \mathbb{C}^k$, representing the receiver front-end noise introduced in $\mathcal{C}_L$, the receive vector $\mathbf{y}_L \in \mathbb{C}^k$ and the linear mapping

$$\mathbf{y}_L = \mathbf{H}_{L+1}\mathbf{s} + \mathbf{n}_L.$$

Here and also in the subsequent equations, the ith elements of the transmit, receive and noise vectors correspond to the ith antenna in the respective network stage.

The time slots $T = \{2, \ldots, L\}$ are used for relaying the signals from cluster to cluster. In time slot T the relay antennas in $\mathcal{C}_{L-T+2}$ transmit scaled versions of the signals received in time slot $T-1$ to $\mathcal{C}_{L-T+1}$. That is, with $l = L - T + 2$ the transmit vector of the lth relay cluster $\mathbf{r}_l \in \mathbb{C}^k$ is computed from the respective receive vector $\mathbf{y}_l \in \mathbb{C}^k$ according to

$$\mathbf{r}_l = \sqrt{\frac{\alpha_l}{k}}\mathbf{y}_l,$$

where $\alpha_l \in \mathbb{R}$ is a cluster specific constant of proportionality specifying the ratio between receive and transmit power. The transmission in time slot T is then described by

$$\mathbf{y}_{l-1} = \mathbf{H}_l\mathbf{r}_l + \mathbf{n}_{l-1}.$$

Here, $\mathbf{H}_l \in \mathbb{C}^{k \times k}$ represents the channel between $\mathcal{C}_l$ and $\mathcal{C}_{l-1}$, $\mathbf{n}_{l-1} \in \mathbb{C}^k$ the receiver front-end noise introduced in $\mathcal{C}_{l-1}$, and $\mathbf{y}_{l-1} \in \mathbb{C}^k$ is the corresponding receive vector. Thus, the signals traverse one hop per time slot. In time slot $T = L+1$, $\mathcal{C}_1$ finally forwards its received signals to the destination. Again, the transmit vector is computed according to $\mathbf{r}_1 = \sqrt{\frac{\alpha_1}{k}}\mathbf{y}_1$. Denoting the matrix representing the

channel between $\mathcal{C}_1$ and the destination by $\mathbf{H}_1$, the vector representing the receiver front-end noise at the destination by $\mathbf{n}_d \in \mathbb{C}^{n_{\mathrm{d}}}$ and the receive vector by $\mathbf{y} \in \mathbb{C}^{n_{\mathrm{d}}}$ this transmission is described by

$$\mathbf{y} = \mathbf{H}_1 \mathbf{r}_1 + \mathbf{n}_{\mathrm{d}}.$$

Putting everything together, the input-output relation of the channel as seen from source to destination antennas over $L+1$ time slots can be written as

$$\mathbf{y} = \sqrt{\frac{\prod_{l=1}^{L} \alpha_l}{k^L}} \mathbf{H}_1 \cdots \mathbf{H}_{L+1} \mathbf{s} + \mathbf{n}_{\mathrm{d}} + \sum_{l=1}^{L} \sqrt{\frac{\prod_{l'=1}^{l} \alpha_{l'}}{k^l}} \mathbf{H}_1 \cdots \mathbf{H}_l \mathbf{n}_l.$$

We model the entries of all noise vectors as zero-mean circular symmetric complex Gaussian random variables of unit-variance that are white both in space and time. The channel matrices are independent and their elements are assumed to be i.i.d. zero-mean random variables of unit-variance. Moreover, we impose the per antenna power constraints $\mathsf{E}[s(i)s(i)^*] = P/n_{\mathrm{s}}$ and $\mathsf{E}[r_l(i)r_l(i)^*] = P/k$ for $l = \{1, \ldots, L\}$. The relay antenna power constraints are fulfilled, if the scaling factors α_i satisfy $\alpha \triangleq \alpha_1 = \ldots = \alpha_L = P/(P+1)$.

IV. Ergodic Capacity & Convergence of Eigenvalues

While full cooperation and the presence of full channel-state information is assumed at the destination antennas, source and relay antennas do not cooperate and also do not possess any channel-state information. Under these assumptions, the ergodic mutual information $I(\mathbf{s};\mathbf{y})$ is maximized, when the entries of $\mathbf{s}$ are zero-mean circularly symmetric complex Gaussian random variables of variance P/n_{s} that are white over both space and time [9]. For this input distribution, $I(\mathbf{s};\mathbf{y})$ is fully characterized by the joint probability distribution of the eigenvalues of the product $\mathbf{C} = \mathbf{R}_s \mathbf{R}_n^{-1}$, where $\mathbf{R}_s \in \mathbb{C}^{n_{\mathrm{d}} \times n_{\mathrm{d}}}$ and $\mathbf{R}_n \in \mathbb{C}^{n_{\mathrm{d}} \times n_{\mathrm{d}}}$ denote the signal and noise covariance matrices at the destination of the multi-hop channel. These covariance

matrices are given by

$$\begin{aligned}\mathbf{R}_s &= \mathsf{E}_{\mathbf{s}}\left[\frac{\alpha^L}{k^L}\mathbf{H}_1\cdots\mathbf{H}_{L+1}\mathbf{s}\mathbf{s}^{\mathrm{H}}\mathbf{H}_{L+1}^{\mathrm{H}}\cdots\mathbf{H}_1^{\mathrm{H}}\right]\\ &= \frac{P\alpha^L}{n_{\mathrm{s}}k^L}\mathbf{H}_1\cdots\mathbf{H}_{L+1}\mathbf{H}_{L+1}^{\mathrm{H}}\cdots\mathbf{H}_1^{\mathrm{H}},\\ \mathbf{R}_n &= \mathsf{E}_{\mathbf{n}_{\mathrm{d}},\mathbf{n}_1,\ldots,\mathbf{n}_L}\left[\left(\mathbf{n}_{\mathrm{d}}+\sum_{l=1}^{L}\sqrt{\frac{\alpha^l}{k^l}}\mathbf{H}_1\cdots\mathbf{H}_l\mathbf{n}_l\right)\left(\mathbf{n}_{\mathrm{d}}+\sum_{l=1}^{L}\sqrt{\frac{\alpha^l}{k^l}}\mathbf{H}_1\cdots\mathbf{H}_l\mathbf{n}_l\right)^{\mathrm{H}}\right]\\ &= \mathbf{I}_{n_{\mathrm{d}}}+\sum_{l=1}^{L}\left(\frac{\alpha}{k}\right)^l\mathbf{H}_1\cdots\mathbf{H}_l\mathbf{H}_l^{\mathrm{H}}\cdots\mathbf{H}_1^{\mathrm{H}}.\end{aligned}$$

We define the the empirical eigenvalue distribution (EED) of the matrix $\mathbf{A}$ as

$$F_{\mathbf{A}}^{(n)}(x) = \frac{1}{n}\sum_{i=1}^{n} 1\{\lambda_i\{\mathbf{A}\} < x\}. \tag{1}$$

With this notation the ergodic capacity of the multi-hop channel in nats per channel use is obtained as

$$\begin{aligned}C &= \frac{1}{L+1}\cdot\mathsf{E}_{\mathbf{C}}\left[\log\det\left(\mathbf{I}_{\mathrm{d}}+\mathbf{C}\right)\right]\\ &= \frac{1}{L+1}\cdot\mathsf{E}_{\mathbf{C}}\left[\sum_{i=1}^{n_{\mathrm{d}}}\log(1+\lambda_i\{\mathbf{C}\})\right]\\ &= \frac{1}{L+1}\cdot\mathsf{E}_{\mathbf{C}}\left[\int_0^\infty\log(1+x)\cdot dF_{\mathbf{C}}^{(n_{\mathrm{d}})}(x)\right].\end{aligned} \tag{2}$$

Note that the pre-log factor $(L+1)^{-1}$ accounting for the use of $L+1$ time slots can be lowered by initiating the next source antenna transmission after $L_0 < L$ time slots. From a practical perspective L_0 needs to be chosen large enough, such that the interference imposed on the previously transmitted signal is negligible. It is important to have this fact in mind whenever we take the limit $L\to\infty$, which formally drives the ergodic capacity to zero. In this paper we are interested in the scaling of the capacity in the number of source and destination antennas. Accordingly, we focus on the case where both these quantities grow large. From [3] we know that $F_{\mathbf{C}}^{(n)}(x)$ converges almost surely (a.s.) to some asymptotic distribution $F_{\mathbf{C}}(x)$, when $n_{\mathrm{s}}\to\infty$, $n_{\mathrm{d}}\to\infty$ and $k\to\infty$, but the ratios $\beta_s = n_{\mathrm{s}}/n_{\mathrm{d}}$ and $\beta_r = k/n_{\mathrm{d}}$ are fixed. Here, we mean by the convergence of an EED $F_{\mathbf{A}}^{(n)}(x)$ to some deterministic function $F_{\mathbf{A}}(x)$ that

$$\mathsf{Pr}\left[\lim_{n\to\infty}\sup_{x\in\mathbb{R}}|F_{\mathbf{A}}^{(n)}(x)-F_{\mathbf{A}}(x)| = 0\right] = 1.$$

We will refer to the density $f_{\mathbf{A}}(x) = \frac{d}{dx} F_{\mathbf{A}}(x)$ as the limiting spectral measure (LSM) subsequently.

Returning to the capacity expression (2), we can infer that for $n_{\mathrm{s}} \to \infty$, $n_{\mathrm{d}} \to \infty$ and $k \to \infty$, and β_s and β_r fixed, C converges to the quantity C_∞ defined as

$$C_\infty \triangleq \frac{1}{L+1} \cdot \int_0^\infty \log(1+x) \cdot f_{\mathbf{C}}(x) \cdot dx.$$

V. Preliminaries From Random Matrix Theory

We briefly repeat some preliminaries from the theory of large random matrices subsequently.

A. *Stieltjes Transform*

We define the Stieltjes transform of some LSM $f(\cdot)$ as

$$G(s) \triangleq \int_{-\infty}^{\infty} \frac{f(x)}{s+x} \cdot dx. \tag{3}$$

We stick to the definition in [5] here, while it is generally more common to define the Stieltjes transform with a minus sign in the denominator. The LSM is uniquely determined by its Stieltjes transform. While it is impossible to find a closed form expression for the LSM of a random matrix in many cases, implicit (polynomial) equations for the corresponding Stieltjes transform can sometimes be obtained. Accordingly, the Stieltjes transform plays a prominent role in large random matrix theory. A transform pair appearing again and again in the course of this paper is the following:

$$f(x) = \delta(x - x_0) \; \Gamma \; G(s) = \frac{1}{s - x_0}. \tag{4}$$

B. *Marčenko-Pastur Law*

The result presented in this paper is valid for the class of random matrices fulfilling the conditions for the Marčenko-Pastur law [10], [11], which we briefly repeat here. Let $\mathbf{X} \in \mathbb{C}^{k_0 \times k_1}$ be a random matrix whose entries are i.i.d. zero-mean distributed and of unit-variance. If both $k_0 \to \infty$ and $k_1 \to \infty$, but $\beta = k_1/k_0$ is kept finite, then the Stieltjes transform of the LSM of $\frac{1}{k_1}\mathbf{X}\mathbf{X}^{\mathrm{H}}$ is given by

$$G_{\mathrm{MP}}^{(\beta)}(s) = \frac{\beta^{-1} - 1 - s \pm \sqrt{s^2 + 2\left(\beta^{-1} + 1\right)s + \left(\beta^{-1} - 1\right)^2}}{2s\beta^{-1}}. \tag{5}$$

The corresponding LSM can be written in closed form.

C. Concatenated Vector Channel

Our result is based on a theorem on the concatenated vector channel proven in [5] using the S-transform [12], [13]:

Theorem 1. *Let* $\mathbf{M}_1 \in \mathbb{C}^{k_0 \times k_1}, \ldots, \mathbf{M}_N \in \mathbb{C}^{k_{N-1} \times k_N}$ *be independent random matrices fulfilling the conditions for the Marčenko-Pastur law, whose elements are of unit-variance, and define* $\beta_n = \frac{k_n}{k_0}$. *Then the Stieltjes transform of the LSM corresponding to* $1/\left(k_1 \cdots k_N\right) \mathbf{M}_1 \cdots \mathbf{M}_N \mathbf{M}_N^{\mathrm{H}} \cdots \mathbf{M}_1^{\mathrm{H}}$ *fulfills the implicit equation*

$$\frac{G(s)}{\beta_N} \prod_{n=0}^{N-1} \frac{sG(s) - 1 + \beta_{n+1}}{\beta_n} + sG(s) = 1. \tag{6}$$

Note that we normalize with respect to k_N rather than with respect to k_0 as done in [5]. The Stieltjes transform $\tilde{G}(s)$ therein relates to $G(s)$ as $G(s) = \beta_N \tilde{G}(\beta_N s)$.

VI. Capacity Scaling

We formalize our result in the subsequent theorem. It is important to note that taking the limits in the LSM of a random matrix means that the dimensions of this matrix are already taken to infinity. For the case at hand that is, we first take the limits $n_{\mathrm{s}} \to \infty$, $n_{\mathrm{d}} \to \infty$ and $k \to \infty$, and then take the limits $L \to \infty$ and $\beta_r \to \infty$. Whenever we take the limits $L \to \infty$ and $\beta_r \to \infty$ in an LSM of some random matrix $\mathbf{A}$ or in the corresponding Stieltjes transform, we denote the asymptotic expressions by

$$f_{\mathbf{A}}^{(\infty)}(x) \triangleq \lim_{L,\beta_r \to \infty} f_{\mathbf{A}}(x) \quad \text{and} \quad G_{\mathbf{A}}^{(\infty)}(s) \triangleq \lim_{L,\beta_r \to \infty} G_{\mathbf{A}}(s).$$

Theorem 2. *Let* $\mathbf{H}_1 \in \mathbb{C}^{n_{\mathrm{d}} \times k}$, $\mathbf{H}_2, \ldots, \mathbf{H}_L \in \mathbb{C}^{k \times k}$ *and* $\mathbf{H}_{L+1} \in \mathbb{C}^{k \times n_{\mathrm{s}}}$ *be independent random matrices with elements of unit-variance fulfilling the conditions needed for the Marčenko-Pastur law and define* $\beta_s \triangleq n_{\mathrm{s}}/n_{\mathrm{d}}$ *and* $\beta_r \triangleq k/n_{\mathrm{d}}$. *Let* snr *be a positive constant. Then, the Stieltjes transform corresponding*

to the LSM of

$$\mathbf{C} = \mathbf{R}_s \cdot \mathbf{R}_n^{-1} = P\alpha^L \cdot \tilde{\mathbf{R}}_s \cdot \left[\sum_{l=0}^{L} \alpha^l \mathbf{R}_{n,l} \right]^{-1},$$

$$\textit{with } P = \frac{1-\alpha^{L+1}}{(1-\alpha)\cdot\alpha^L} \cdot \mathsf{snr},$$

$$\alpha = \frac{P}{1+P},$$

$$\tilde{\mathbf{R}}_s \triangleq \frac{1}{n_{\mathrm{s}} k^L} \mathbf{H}_1 \cdots \mathbf{H}_{L+1} \mathbf{H}_{L+1}^{\mathrm{H}} \cdots \mathbf{H}_1^{\mathrm{H}},$$

$$\mathbf{R}_{n,l} \triangleq \begin{cases} \mathbf{I}_{n_{\mathrm{d}}} & , \text{if } l = 0, \\ \frac{1}{k^l} \mathbf{H}_1 \cdots \mathbf{H}_l \mathbf{H}_l^{\mathrm{H}} \cdots \mathbf{H}_1^{\mathrm{H}} & , \text{else.} \end{cases}$$

in the limits $\beta_r \to \infty$, $L \to \infty$, but β_s fixed, converges to

$$G_{\mathbf{C}}^{(\infty)}(s) = \begin{cases} \frac{1}{\mathsf{snr}} G_{\mathrm{MP}}^{(\beta_s)}\left(\frac{s}{\mathsf{snr}}\right), \text{if } L \in \mathcal{O}\left(\beta_r^{1-\varepsilon}\right), \varepsilon > 0, \\ s^{-1} \qquad\qquad\quad , \text{if } L \in \Omega(\beta_r^{1+\varepsilon}), \varepsilon > 0. \end{cases} \tag{7}$$

Furthermore, if $L \in \Theta(\beta_r)$ the Stieltjes transform converges to some $G_{\mathbf{C}}^{(\infty)}(s) \neq s^{-1}$ in this limit.

We have introduced the parameters P and snr such that they correspond to the average transmit power at the source and the average SNR at the destination. The case $G_{\mathbf{C}}^{(\infty)}(s) = \frac{1}{\mathsf{snr}} G_{\mathrm{MP}}^{(\beta_s)}\left(\frac{s}{\mathsf{snr}}\right)$ corresponds to a point-to-point MIMO channel scenario and thus generalizes [7] and [3] in the sense that it gives a condition on how fast the number of relays per layer needs to grow with the number of hops in order to approach the Marčenko-Pastur law for increasing L. The case $G_{\mathbf{C}}^{(\infty)}(s) = s^{-1}$ can be seen as a generalization of Theorem 4 in [5], which states that almost all eigenvalues vanish, i.e., $f_{\mathbf{C}}^{(\infty)}(x) = \delta(x)$ a.s., if the aspect ratios β_r remain finite. The case of linear scaling of β_r in L constitutes the threshold between the previous two regimes, but still suffices in order to sustain a non-zero fraction of the spatial degrees of freedom. In summary, the theorem thus states that the capacity scales linearly in $\min\{n_{\mathrm{s}}, n_{\mathrm{d}}\}$ as long as k scales at least linearly in both L and $\min\{n_{\mathrm{s}}, n_{\mathrm{d}}\}$ and the SNR at the destination is kept constant.

VII. Asymptotic Analysis

This section provides the proof of Theorem 2. We start out with the lemma below, which will allow for inferring a corollary to Theorem 1. This corollary is the key to the proof of Theorem 2 in this section. In

order not to interrupt the logical flow of the section two rather technical lemmas required for the proof of Theorem 2, are stated and proven in the Appendix of this paper.

Lemma 1. *For any* $\varepsilon > 0$*, some function* $g : \mathbb{R} \longrightarrow \mathbb{R}_+$*, and a positive constant* c

$$\lim_{\kappa\to\infty} \left(\frac{c}{\kappa}+1\right)^{g(\kappa)} = \begin{cases} 1, & \text{if} \quad g(\kappa) \in \mathcal{O}(\kappa^{1-\varepsilon}), \\ \infty, & \text{if} \quad g(\kappa) \in \Omega(\kappa^{1+\varepsilon}). \end{cases}$$

Furthermore, if $g(\kappa) \in \Theta(\kappa)$ *there exist constants* $M_2 \geq M_1 > 0$*, such that*

$$\liminf_{\kappa\to\infty} \left(\frac{c}{\kappa}+1\right)^{g(\kappa)} = e^{cM_1},$$
$$\limsup_{\kappa\to\infty} \left(\frac{c}{\kappa}+1\right)^{g(\kappa)} = e^{cM_2}.$$

Proof. The lemma follows from the fact that the limit can be taken inside a continuous function, which allows as to write

$$\lim_{\kappa\to\infty} \left(\frac{c}{\kappa}+1\right)^{g(\kappa)} = \exp\left(\lim_{\kappa\to\infty} g(\kappa)\cdot\log\left(\frac{c}{\kappa}+1\right)\right), \tag{8}$$

and the rule of Bernoulli-l'Hospital applied to $g(\kappa) = M\kappa^{\gamma}$, where M and γ are positive constants:

$$\lim_{\kappa\to\infty} M\kappa^{\gamma}\cdot\log\left(\frac{c}{\kappa}+1\right) = \frac{cM\kappa^{\gamma}}{\gamma\,(c+\kappa)}. \tag{9}$$

If $g(\kappa) \in \mathcal{O}(\kappa^{1-\varepsilon})$, by definition there exists some $M > 0$, such that the exponent in (8) can be upper-bounded by

$$\lim_{\kappa\to\infty} g(\kappa)\cdot\log\left(\frac{c}{\kappa}+1\right) \leq \lim_{\kappa\to\infty} M\kappa^{1-\varepsilon}\cdot\log\left(\frac{c}{\kappa}+1\right). \tag{10}$$

Evaluating (9) for $\gamma = 1-\varepsilon$ renders this upper bound zero, which establishes that also the left hand side (LHS) of (10) becomes zero and (8) evaluates to one in this case.

Analogously, if $g(\kappa) \in \Omega(\kappa^{1+\varepsilon})$, there exists some $M > 0$, such that the exponent in (8) can be lower bounded according to

$$\lim_{\kappa\to\infty} g(\kappa)\cdot\log\left(\frac{c}{\kappa}+1\right) \geq \lim_{\kappa\to\infty} M\kappa^{1+\varepsilon}\cdot\log\left(\frac{c}{\kappa}+1\right). \tag{11}$$

Evaluating (9) for $\gamma = 1 + \varepsilon$ renders the upper bound infinite, which establishes that both the LHS of (11) and (8) grow without bound in this case.

Finally, if $g(\kappa) \in \Theta(\kappa)$, there exist constants M_1 and M_2, fulfilling $M_2 \geq M_1 > 0$, such that according to (9) evaluated for $\gamma = 1$ the exponent in (8) is sandwiched between

$$\begin{aligned} cM_1 &= \lim_{\kappa\to\infty} M_1\kappa \cdot \log\left(\frac{c}{\kappa}+1\right) \\ &\leq \lim_{\kappa\to\infty} g(\kappa) \cdot \log\left(\frac{c}{\kappa}+1\right) \\ &\leq \lim_{\kappa\to\infty} M_2\kappa \cdot \log\left(\frac{c}{\kappa}+1\right) = cM_2, \end{aligned}$$

which establishes the second part of the lemma. ■

Corollary 1 of Theorem 1. *With the notation and assumptions from Theorem 2 the Stieltjes transform of the LSM corresponding to* $\tilde{\mathbf{R}}_s$ *in the limit* $k \to \infty$ *and* $L \to \infty$ *converges to*

$$G_{\tilde{\mathbf{R}}_s}^{(\infty)}(s) = \begin{cases} G_{\mathrm{MP}}^{(\beta_s)}(s), & \text{if} \quad L \in \mathcal{O}(\beta_r^{1-\varepsilon}), \\ s^{-1}, & \text{if} \quad L \in \Omega(\beta_r^{1+\varepsilon}). \end{cases} \tag{12}$$

Also, if $L \in \Theta(\beta_r)$ *the LSM of* $\tilde{\mathbf{R}}_s$ *converges to a distribution corresponding to a Stieltjes transform* $G_{\tilde{\mathbf{R}}_s}^{(\infty)}(s) \neq s^{-1}$ *in this limit.*

Furthermore, if $L \in \mathcal{O}(\beta_r^{1-\varepsilon})$ *the Stieltjes transforms of the LSMs corresponding to the* $\mathbf{R}_{n,l}$*'s, for* $l \in \{1, \ldots, L\}$*, in the limit* $k \to \infty$ *and* $L \to \infty$ *converge to*

$$G_{\mathbf{R}_{n,l}}^{(\infty)}(s) = (s-1)^{-1}. \tag{13}$$

The part of the corollary related to the $\mathbf{R}_{n,l}$'s is actually stated more generally than needed in this paper. In fact, we are only going to use that for some fixed positive integer $L_\mathrm{t} < L$, the LSMs of the $\mathbf{R}_{n,l}$'s, where $l \in \{1, \ldots, L_\mathrm{t}\}$, are given by a Dirac delta at one, i.e., have the Stieltjes transform (13), when $\beta_r \to \infty$, independently of the scaling of β_r in L. This is trivially guaranteed by the corollary, since when L_t is constant, $L_\mathrm{t} \in \mathcal{O}(\beta_r^{1-\varepsilon})$ is fulfilled naturally as $\beta_r \to \infty$.

Proof of Corollary We treat (12) first. An implicit equation for the Stieltjes transform of the LSM corresponding to $\tilde{\mathbf{R}}_s$ is given by (6), where we set $N = L + 1$, $\beta_0 = 1$, $\beta_n = \beta_r$ for $n \in \{1, \ldots N-1\}$,

and $\beta_N = \beta_s$ according to our notation:

$$\frac{G_{\tilde{\mathbf{R}}_s}}{\beta_s} \cdot \frac{sG_{\tilde{\mathbf{R}}_s} - 1 + \beta_s}{\beta_r} \cdot \overbrace{\left(\frac{sG_{\tilde{\mathbf{R}}_s} - 1 + \beta_r}{\beta_r}\right)^{L-1}}^{\Psi_s(\beta_r, L)} \cdot \left(sG_{\tilde{\mathbf{R}}_s} - 1 + \beta_r\right) + sG_{\tilde{\mathbf{R}}_s} = 1. \tag{14}$$

We apply Lemma 1 to $\Psi_s(\beta_r, L)$, where we identify β_r with κ and L with $g(\kappa)$. In the limits $\beta_r \to \infty$, $L \to \infty$ and $L \in \mathcal{O}(\beta_r^{1-\varepsilon})$ this yields $\Psi_s(\beta_r, L) \to 1$. Accordingly, (14) simplifies to the quadratic equation

$$\beta_s^{-1} s G_{\tilde{\mathbf{R}}_s}^{(\infty)^2}(s) + \left(s + 1 - \beta_s^{-1}\right) G_{\tilde{\mathbf{R}}_s}^{(\infty)} = 1. \tag{15}$$

in this limit. The solution to (15) is the Stieltjes transform of the Marčenko-Pastur law (5) with $\beta = \beta_s$. If $L \in \Omega(\beta_r^{1+\varepsilon})$ we know from Lemma 1 that $\Psi_s(\beta_r, L)$ grows without bounds for $L \to \infty$. The numbers of factors in the LHS of (6) grows with L in this case. Theorem 4 in reference [5] states that for $N \to \infty$ the Stieltjes transform in (6) converges to $G_{\tilde{\mathbf{R}}_s}^{(\infty)} = s^{-1}$, if the β_n are uniformly bounded. In fact, the conditions needed for this theorem when $\tilde{\beta} \triangleq \beta_1 = \ldots = \beta_{N-1}$, can be relaxed to $N \in \Omega(\tilde{\beta}^{1+\varepsilon})$, while the proof in [5] remains valid. Accordingly, the second case in (12) follows immediately.

For the case $L \in \Theta(\beta_r)$ we know from Lemma 1 that $\Psi_s(\beta_r, L) = e^{d(sG_{\tilde{\mathbf{R}}_s}^{(\infty)}(s) - 1)}$ in the limit of interest for some $d > 0$. Thus (6) simplifies to

$$G_{\tilde{\mathbf{R}}_s}^{(\infty)}(s) \left(e^{d(sG_{\tilde{\mathbf{R}}_s}^{(\infty)}(s) - 1)} \left(\beta_s^{-1} s G_{\tilde{\mathbf{R}}_s}^{(\infty)} + 1 - \beta_s^{-1}\right) + s\right) = 1. \tag{16}$$

There exists no closed form solution to this implicit equation. However, it is easily verified that $G_{\tilde{\mathbf{R}}_s}^{(\infty)} = s^{-1}$ does not satisfy (16).

For (13), we obtain the equation for the Stieltjes transform corresponding to the LSM of $\mathbf{R}_{n,l}$, where $l \in \{1, \ldots, L\}$, from (6) with $N = l$, $\beta_0 = 1$ and $\beta_n = \beta_r$ for $n \in \{1, \ldots, l\}$ as

$$\frac{G_{\mathbf{R}_{n,l}}(s)}{\beta_r} \cdot \overbrace{\left(\frac{sG_{\mathbf{R}_{n,l}}(s) - 1 + \beta_r}{\beta_r}\right)^{l-1}}^{\Psi_n(\beta_r, l)} \cdot \left(sG_{\mathbf{R}_{n,l}}(s) - 1 + \beta_r\right) + sG_{\mathbf{R}_{n,l}}(s) = 1. \tag{17}$$

Let's consider the case $l = L$ first. If $L \in \mathcal{O}\left(\beta_r^{1-\varepsilon}\right)$, $\Psi_n(\beta_r, l)$ converges to one in the limit $k \to \infty$ and $L \to \infty$ by Lemma 1. Therefore, in this limit (17) simplifies to

$$G_{\mathbf{R}_{n,L}}^{(\infty)}(s) + sG_{\mathbf{R}_{n,L}}^{(\infty)}(s) = 1. \tag{18}$$

The solution to (18) is given by $G^{(\infty)}_{\mathbf{R}_{n,L}}(s) = (s+1)^{-1}$. The same is trivially true for all $\mathbf{R}_{n,l}$ with $l < L$, since whenever $L \in \mathcal{O}\left(\beta_r^{1-\varepsilon}\right)$, this implies that also $l \in \mathcal{O}(\beta_r^{1-\varepsilon})$. ■

We are now ready to prove Theorem 2. Besides the above corollary, we use the Lemmas 2 & 3, which are stated and proven in the Appendix of this paper.

Proof of Theorem 2. We go through the different scaling behaviors of β_r with L, subsequently.

Cases $L \in \mathcal{O}\left(\beta_r^{1-\varepsilon}\right)$ and $L \in \Theta\left(\beta_r\right)$: Firstly, we show for *both* these cases that the LSM of $\mathbf{R}_n$ takes on the shape of a Dirac delta at some positive constant in the limit of interest. The proof is based on a truncation of the relay chain between the stages L_{t} and $L_t - 1$. By choosing L_{t} large enough, we can achieve that the accumulated noise power originating from the relay stages $L_t, \ldots, L$ is sufficiently attenuated before it reaches the destination. More specifically, we claim that for any $\varepsilon > 0$ there exist positive integers $L_{\mathrm{t}}^{(1)}$ and $n_0^{(1)}(L, L_{\mathrm{t}})$[1], such that for all $n \geq n_0^{(1)}(L, L_{\mathrm{t}})$, for all $L_{\mathrm{t}} > L_{\mathrm{t}}^{(1)}$ and L arbitrarily large a.s.[2]

$$\frac{1}{n_{\mathrm{d}}} \left\| \sum_{l=L_{\mathrm{t}}+1}^{L} \left(\frac{\alpha}{k}\right)^{l} \mathbf{H}_1 \cdots \mathbf{H}_l \mathbf{H}_l^{\mathrm{H}} \cdots \mathbf{H}_1^{\mathrm{H}} \right\|_{\mathrm{Tr}} < \frac{\varepsilon}{3}. \tag{19}$$

We prove this by the following chain of inequalities:

$$\begin{aligned}
&\frac{1}{n_{\mathrm{d}}} \left\| \sum_{l=L_{\mathrm{t}}}^{L} \left(\frac{\alpha}{k}\right)^{l} \mathbf{H}_1 \cdots \mathbf{H}_l \mathbf{H}_l^{\mathrm{H}} \cdots \mathbf{H}_1^{\mathrm{H}} \right\|_{\mathrm{Tr}} \\
&\quad \leq \frac{1}{n_{\mathrm{d}}} \sum_{l=L_{\mathrm{t}}}^{L} \alpha^{l} \left\| \frac{1}{k^{l}} \mathbf{H}_1 \cdots \mathbf{H}_l \mathbf{H}_l^{\mathrm{H}} \cdots \mathbf{H}_1^{\mathrm{H}} \right\|_{\mathrm{Tr}} \\
&\quad \leq \max_{l'=L_t,\ldots,L} \left\{ \frac{1}{n_{\mathrm{d}}} \left\| \frac{1}{k^{l'}} \mathbf{H}_1 \cdots \mathbf{H}_{l'} \mathbf{H}_{l'}^{\mathrm{H}} \cdots \mathbf{H}_1^{\mathrm{H}} \right\|_{\mathrm{Tr}} \right\} \cdot \sum_{l=L_{\mathrm{t}}+1}^{\infty} \alpha^{l} \\
&\quad = \max_{l'=L_t,\ldots,L} \left\{ \frac{1}{n_{\mathrm{d}}} \left\| \frac{1}{k^{l'}} \mathbf{H}_1 \cdots \mathbf{H}_{l'} \mathbf{H}_{l'}^{\mathrm{H}} \cdots \mathbf{H}_1^{\mathrm{H}} \right\|_{\mathrm{Tr}} \right\} \cdot \frac{\alpha^{L_{\mathrm{t}}}}{1-\alpha}.
\end{aligned} \tag{20}$$

In the first step we applied the triangle inequality and used the homogeneity of the Trace norm. In the second step we upper-bounded the coefficients of the α^{l}'s by the maximum coefficient. Afterwards, we let the number of summands go to infinity, which strictly increases the term, since all added summands

[1] We write $n_0^{(1)}(L, L_{\mathrm{t}})$ in order to emphasize that $n_0(1)$ is a function of L and L_{t}

[2] The Trace norm of the matrix $\mathbf{A} \in \mathbb{C}^{n\times n}$ is defined as $\|\mathbf{A}\|_{\mathrm{Tr}} = \sum_{i=1}^{n} \lambda_i\{\mathbf{A}\}$.

are positive. A standard identity for geometric series allows for eliminating the sum. *All* arguments of the $\max\{\cdot\}$ function in (20) converge to one a.s.. This follows immediately from Theorem 3 in [5]. Therefore, we can choose an $n_0^{(1)}(L, L_\mathrm{t})$ large enough, such that even the maximum of the $L - L_\mathrm{t} + 1$ terms is arbitrarily close to one a.s. for all $n \geq n_0^{(1)}(L, L_\mathrm{t})$. In particular, if $L_\mathrm{t} > \log_\alpha((1-\alpha) \cdot \varepsilon/3)$, we can make $n_0^{(1)}(L, L_\mathrm{t})$ large enough, such that (19) is fulfilled a.s. for all $n \geq n_0^{(1)}(L, L_\mathrm{t})$.

In a next step we can choose L (and thus β_r) large enough such that the accumulated noise power originating from the relays $1, \ldots, L_\mathrm{t} - 1$ becomes sufficiently white. This means, for the fixed L_t and the same ε as defined above there exist $L_0 > L_\mathrm{t}$ and $n_0^{(2)}(L, L_\mathrm{t})$, such that a.s. for all $L \geq L_0$ and for all $n \geq n_0^{(2)}(L, L_\mathrm{t})$

$$\frac{1}{n_\mathrm{d}} \left\| \frac{1-\alpha^{L_\mathrm{t}}}{1-\alpha} \mathbf{I}_{n_\mathrm{d}} - \sum_{l=0}^{L_\mathrm{t}-1} \left(\frac{\alpha}{k}\right)^l \mathbf{H}_1 \cdots \mathbf{H}_l \mathbf{H}_l^\mathrm{H} \cdots \mathbf{H}_1^\mathrm{H} \right\|_\mathrm{Tr} < \frac{\varepsilon}{3}.$$

The proof is similar to the one above:

$$\begin{aligned}
&\frac{1}{n_\mathrm{d}} \left\| \frac{1-\alpha^{L_\mathrm{t}}}{1-\alpha} \mathbf{I}_{n_\mathrm{d}} - \sum_{l=0}^{L_\mathrm{t}-1} \left(\frac{\alpha}{k}\right)^l \mathbf{H}_1 \cdots \mathbf{H}_l \mathbf{H}_l^\mathrm{H} \cdots \mathbf{H}_1^\mathrm{H} \right\|_\mathrm{Tr} \\
&= \frac{1}{n_\mathrm{d}} \left\| \sum_{l=0}^{L_\mathrm{t}-1} \alpha^l \cdot \left(\mathbf{I}_{n_\mathrm{d}} - \frac{1}{k^l} \mathbf{H}_1 \cdots \mathbf{H}_l \mathbf{H}_l^\mathrm{H} \cdots \mathbf{H}_1^\mathrm{H} \right) \right\|_\mathrm{Tr} \\
&\leq \frac{1}{n_\mathrm{d}} \sum_{l=0}^{L_\mathrm{t}-1} \alpha^l \left\| \mathbf{I}_{n_\mathrm{d}} - \frac{1}{k^l} \mathbf{H}_1 \cdots \mathbf{H}_l \mathbf{H}_l^\mathrm{H} \cdots \mathbf{H}_1^\mathrm{H} \right\|_\mathrm{Tr} \\
&\leq \max_{l' \in \{1,\ldots,L_\mathrm{t}-1\}} \left\{ \frac{1}{n_\mathrm{d}} \left\| \mathbf{I}_{n_\mathrm{d}} - \frac{1}{k^l} \mathbf{H}_1 \cdots \mathbf{H}_l \mathbf{H}_l^\mathrm{H} \cdots \mathbf{H}_1^\mathrm{H} \right\|_\mathrm{Tr} \right\} \cdot \sum_{l=0}^{\infty} \alpha^l \\
&= \max_{l' \in \{1,\ldots,L_\mathrm{t}-1\}} \left\{ \frac{1}{n_\mathrm{d}} \left\| \mathbf{I}_{n_\mathrm{d}} - \frac{1}{k^l} \mathbf{H}_1 \cdots \mathbf{H}_l \mathbf{H}_l^\mathrm{H} \cdots \mathbf{H}_1^\mathrm{H} \right\|_\mathrm{Tr} \right\} \cdot \frac{1}{1-\alpha}.
\end{aligned} \tag{21}$$

Again, the first identity is a standard identity for a geometric series. In this case the convergence of the arguments of the $\max(\cdot)$ function is guaranteed by Corollary 1 and Lemma 2 (refer to Appendix): Corollary 1 tells us that the LSMs of all the $\frac{1}{k^l}\mathbf{H}_1 \cdots \mathbf{H}_l \mathbf{H}_l^\mathrm{H} \cdots \mathbf{H}_1^\mathrm{H}$'s, where $l \in \{1, \ldots, L_\mathrm{t} - 1\}$, converge to a Dirac at one in the limit under consideration. Knowing this, Lemma 2 guarantees us that the respective Trace norms go to zero. Thus, there exist L_0 and $n_0^{(2)}(L, L_\mathrm{t})$, such that a.s. the maximum of the $L_\mathrm{t} - 1$ terms is small enough to make (21) smaller than $\varepsilon/3$ for all $L > L_0$ and $n \geq n_0^{(2)}(L, L_\mathrm{t})$.

With the choices $L_{\mathrm{t}} = \max\{L_{\mathrm{t}}^{(1)}, \log_\alpha((1-\alpha)\cdot\varepsilon/3)\}$ and $n_0(L, L_{\mathrm{t}}) = \max\{n_0^{(1)}(L, L_{\mathrm{t}}), n_0^{(2)}(L, L_{\mathrm{t}})\}$, we can finally conclude by the triangle inequality that for all $L > L_0$ and $n \geq n_0(L, L_{\mathrm{t}})$ a.s.

$$\begin{aligned}
\frac{1}{n_{\mathrm{d}}}\left\|\frac{1-\alpha^{L+1}}{1-\alpha}\cdot\mathbf{I}_{n_{\mathrm{d}}} - \mathbf{R}_n\right\|_{\mathrm{Tr}} &= \frac{1}{n_{\mathrm{d}}}\left\|\frac{1-\alpha^{L+1}}{1-\alpha}\cdot\mathbf{I}_{n_{\mathrm{d}}} - \sum_{l=0}^{L}\alpha^l\mathbf{R}_{n,l}\right\|_{\mathrm{Tr}} \\
&= \frac{1}{n_{\mathrm{d}}}\left\|\frac{1-\alpha^{L+1}}{1-\alpha}\cdot\mathbf{I}_{n_{\mathrm{d}}} - \sum_{l=0}^{L}\left(\frac{\alpha}{k}\right)^l\mathbf{H}_1\cdots\mathbf{H}_l\mathbf{H}_l^{\mathrm{H}}\cdots\mathbf{H}_1^{\mathrm{H}}\right\|_{\mathrm{Tr}} \\
&= \frac{1}{n_{\mathrm{d}}}\left\|\frac{\alpha^{L_{\mathrm{t}}}-\alpha^{L}}{1-\alpha}\cdot\mathbf{I}_{n_{\mathrm{d}}} + \frac{1-\alpha^{L_{\mathrm{t}}}}{1-\alpha}\cdot\mathbf{I}_{n_{\mathrm{d}}} - \sum_{l=0}^{L_{\mathrm{t}}-1}\left(\frac{\alpha}{k}\right)^l\mathbf{H}_1\cdots\mathbf{H}_l\mathbf{H}_l^{\mathrm{H}}\cdots\mathbf{H}_1^{\mathrm{H}} - \sum_{l=L_{\mathrm{t}}}^{L}\left(\frac{\alpha}{k}\right)^l\mathbf{H}_1\cdots\mathbf{H}_l\mathbf{H}_l^{\mathrm{H}}\cdots\mathbf{H}_1^{\mathrm{H}}\right\|_{\mathrm{Tr}} \\
&\leq \frac{1}{n_{\mathrm{d}}}\left\|\frac{\alpha^{L_{\mathrm{t}}}}{1-\alpha}\cdot\mathbf{I}_{n_{\mathrm{d}}}\right\|_{\mathrm{Tr}} + \frac{1}{n_{\mathrm{d}}}\left\|\frac{1-\alpha^{L_{\mathrm{t}}}}{1-\alpha}\cdot\mathbf{I}_{n_{\mathrm{d}}} - \sum_{l=0}^{L_{\mathrm{t}}-1}\left(\frac{\alpha}{k}\right)^l\mathbf{H}_1\cdots\mathbf{H}_l\mathbf{H}_l^{\mathrm{H}}\cdots\mathbf{H}_1^{\mathrm{H}}\right\|_{\mathrm{Tr}} \\
&\quad + \frac{1}{n_{\mathrm{d}}}\left\|\sum_{l=L_{\mathrm{t}}}^{L}\left(\frac{\alpha}{k}\right)^l\mathbf{H}_1\cdots\mathbf{H}_l\mathbf{H}_l^{\mathrm{H}}\cdots\mathbf{H}_1^{\mathrm{H}}\right\|_{\mathrm{Tr}} < \frac{\varepsilon}{3}+\frac{\varepsilon}{3}+\frac{\varepsilon}{3} = \varepsilon.
\end{aligned}$$

Here, we decomposed the terms in a way that allowed us to obtain a sum of precisely the expressions we had proven to converge to zero before. Besides standard steps we used the fact that that $|\alpha^{L_{\mathrm{t}}} - \alpha^L| < |\alpha^{L_{\mathrm{t}}}|$, since $L > L_t$ and $\alpha < 1$. By Lemma 2 we have established that the LSM of $\mathbf{R}_n$ converges to

$$f_{\mathbf{R}_n}^{(\infty)}(x) = \delta\left(x - \frac{1-\alpha^{L+1}}{1-\alpha}\right). \tag{22}$$

Note that the fact that *almost all* eigenvalues of the single $\mathbf{R}_{n,l}$'s are *arbitrarily close* to one each, does not immediately imply that this is also the case for a weighted sum of these matrices. This is due to the fact that for the matrix series $\mathbf{R}_{n,l}$, where $l \in \{1,\dots L_{\mathrm{t}}\}$, the identity $\lambda_k\{\sum_{l=0}^{L}\mathbf{R}_{n,l}\} = \sum_{l=0}^{L}\lambda_k\{\mathbf{R}_{n,k}\}$ is fulfilled, if and only if *all* these eigenvalues are *exactly* equal to one. This easy attempt of proving (22) must therefore fail. Also note that the $\mathbf{R}_{n,l}$'s are *not* asymptotically free, which prohibits arguing based on the respective R-transforms [13].

Since the eigenvalues of the corresponding inverse are the inverse eigenvalues, i.e., $\lambda_k\{\mathbf{R}_n^{-1}\} = \lambda_k^{-1}\{\mathbf{R}_n\}$, we conclude that the LSM of $\mathbf{R}_n^{-1}$ is given by $f_{\mathbf{R}_n^{-1}}^{(\infty)}(x) = \delta(x - (1+\alpha)/(1-\alpha^{L+1}))$. Thus, by Lemma 3 (refer to Appendix) and the respective variable transformation applied to $f_{\mathbf{R}_n^{-1}}^{(\infty)}(\cdot)$ the EED of $\mathbf{C} = \mathsf{snr}\cdot\tilde{\mathbf{R}}_s\mathbf{R}_n^{-1}$ coincides with the EED of $\mathsf{snr}\cdot\tilde{\mathbf{R}}_s$, i.e.,

$$f_{\mathbf{C}}^{(\infty)}(x) = \frac{1}{\mathsf{snr}}f_{\tilde{\mathbf{R}}_s}\left(\frac{x}{\mathsf{snr}}\right) \quad \text{and} \quad G_{\mathbf{C}}^{(\infty)}(s) = \frac{1}{\mathsf{snr}}G_{\tilde{\mathbf{R}}_s}\left(\frac{s}{\mathsf{snr}}\right).$$

By Corollary 1 the LSM of $\tilde{\mathbf{R}}_s$ is given by the Marčenko-Pastur law, in the case that $L \in \mathcal{O}(\beta_r^{1-\varepsilon})$, which establishes the first case. In the case $L \in \Theta(\beta_r)$ a non-zero fraction of the eigenvalues of $\tilde{\mathbf{R}}_s$ remains non-zero as $L \to \infty$ by Corollary 1, i.e., $G_{\tilde{\mathbf{R}}_s}(s) = G_{\mathbf{C}}(s) \neq s^1$.

Case $L \in \Omega(\beta_r^{1+\varepsilon})$**:** This case follows immediately by Corollary 1. Since asymptotically almost all eigenvalues of $\tilde{\mathbf{R}}_s$ vanish, also almost all eigenvalues of $\mathbf{C} = \mathbf{R}_s\mathbf{R}_n^{-1}$ need to approach zero. We rely on the reader's intuition here, that noise cannot recover degrees of freedom. A formal proof goes along the lines of the proof of Lemma 3, where $\mathbf{A}$ is identified with $\mathbf{R}_n^{-1}$ and $\mathbf{B}$ with $\mathbf{R}_s$. In the end one can show that the Shannon transforms of the LSMs $f_{\mathbf{R}_s\mathbf{R}_n^{-1}}(\cdot)$ and $f_{\mathbf{R}_s}(\cdot)$ coincide at a Dirac delta at zero in the limit $L \to \infty$ and $\beta_r \to \infty$. ■

VIII. Simulation Results

The above results provide no evidence about speed of convergence. Since speed of convergence results are generally hard to obtain in a large matrix dimensions analysis, we resort to a numerical demonstration for this purpose. In doing so, we specify the distribution of the elements of $\mathbf{H}_1, \ldots, \mathbf{H}_{L+1}$ as circularly symmetric complex Gaussian with zero-mean and unit-variance. Recall that all assumptions imposed on this distribution in Theorem 2 were just related to its first and second moments. We fix the number of source and destination antennas to $n \triangleq n_{\mathrm{s}} = n_{\mathrm{d}} = 10$ and plot the normalized ergodic capacity $C_0 = (L+1) \cdot C/n$ as obtained through Monte Carlo simulations versus the number of relay clusters, L, for an average SNR of 10 dB at the destination. The number of relays per cluster evolves with the number of clusters according to $k = L^\gamma$, where we vary γ between 0 and 3. In principle, one could use the recursive formula for the Stieltjes transform of $\mathbf{C}$ obtained in [3] rather than Monte Carlo simulations for generating these plots. However, the respective evaluations are handy for small L only.

Fig. 2 shows the case of linear and faster scaling of k in n. For $\gamma = 1$ the curve flattens out quickly, and converges to some constant which is smaller than the normalized point-to-point MIMO capacity, but non-zero, as expected. Furthermore, we observe that the point-to-point limit is approached the sooner the bigger γ is chosen. For $\gamma = 3$ this is the case after less than 10 hops already. Fig. 3 shows the case of

less than linear scaling of k in n. While C_0 decreases rather rapidly for constant relay numbers ($\gamma = 0$), we observe that already a moderate growth of k with L slows down the capacity decay significantly. For $\gamma = 0.5$ an almost threefold capacity gain over the $\gamma = 0$ case is achieved for $L = 16$. For $\gamma = 0.75$ the decay is tolerable even for very large L. Note that for $L = 81$ and $\gamma = 0.75$ only 27 relays per layer are used in contrast to the 81 relays needed for linear scaling.

IX. Conclusion

We have given a criterion how the number of relays per stage needs to be increased with the number of hops in order to sustain a non-zero fraction of the spatial degrees of freedom in a MIMO amplify-and-forward multi-hop network, i.e., linear capacity scaling in $\min\{n_{\mathrm{s}}, n_{\mathrm{d}}\}$. The necessary and sufficient condition is an *at least linear* scaling of the relays per stage in the number of hops.

X. Acknowledgement

We gratefully acknowledge the very helpful comments by Veniamin I. Morgenshtern.

Appendix

Lemma 2. *A random matrix* $\mathbf{A} \in \mathbb{C}^{n\times n}$ *fulfilling* $\lim_{n\to\infty} n^{-1}\mathrm{Tr}(\mathbf{A}) = 1$ *converges to the identity matrix a.s. in the sense that*

$$\lim_{n\to\infty} \frac{1}{n} \|\mathbf{I}_n - \mathbf{A}\|_{\mathrm{Tr}} = 0,$$

if and only if its EED $F^n_{\mathbf{A}}(x)$ *converges to* $\sigma(x-1)$*, i.e., a.s.*

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}} |F^n_{\mathbf{A}}(x) - \sigma(x-1)| = 0.$$

Proof. The lemma follows by a the following identities:

$$
\begin{aligned}
&\lim_{n\to\infty}\frac{1}{n}\|\mathbf{I}_n-\mathbf{A}\|_{\mathrm{Tr}} = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}|1-\lambda_i\{\mathbf{A}\}| \\
&= \lim_{n\to\infty}\frac{1}{n}\sum_{i:\lambda_i\{\mathbf{A}\}\le 1}(1-\lambda_i\{\mathbf{A}\})+\frac{1}{n}\sum_{i:\lambda_i\{\mathbf{A}\}>1}(\lambda_i\{\mathbf{A}\}-1) \qquad (23)\\
&= \lim_{n\to\infty}\int_0^1 |F_{\mathbf{A}}^{(n)}(x)|\cdot dx+\int_1^\infty |F_{\mathbf{A}}^{(n)}(x)-1|\cdot dx \\
&= \lim_{n\to\infty}\int_0^\infty |F_{\mathbf{A}}^{(n)}(x)-\sigma(x-1)|\cdot dx = 0 \qquad (24)\\
\iff &\lim_{n\to\infty}\sup_{x\in\mathbb{R}}|F_{\mathbf{A}}^{(n)}(x)-\sigma(x-1)| = 0. \qquad (25)
\end{aligned}
$$

In (23) we arrange the individual summands such that they can be related to the EED of $\mathbf{A}$. The equivalence of the norms in (24) and (25) is established as follows: For the forward direction consider $\epsilon(x) \triangleq |F^{(n)}(x)-\sigma(x-1)|$ for $x\in[0,1)$, i.e., $\epsilon(x)=|F^{(n)}(x)|$. Choose any $\Delta\in[-1,0)$. Since $\epsilon(x)$ is monotonically increasing on the interval of interest, we can write

$$
\int_{1-\Delta}^{1}|F_{\mathbf{A}}^{(n)}(x)-\sigma(x-1)|\cdot dx > |\Delta|\cdot\epsilon(1+\Delta).
$$

Thus, if $\epsilon(1+\Delta)$ does not go to zero for all Δ, the integral norm cannot go to zero. The same reasoning can be applied for the interval $\Delta\in[1,\infty)$.

For the backward part we break the integration in (24) into two parts. The first integral is from zero to some constant $d>1$. This part is a Riemann integral over a function that converges uniformly by (25). It goes to zero by taking the limit inside the integral. The second part of the integral is from d to ∞. Here, the limit cannot be taken inside the integral in general. However, we can write

$$
\lim_{n\to\infty}\int_d^\infty 1-F_{\mathbf{A}}^{(n)}(x)\cdot dx = \lim_{n\to\infty}\int_0^\infty 1-F_{\mathbf{A}}^{(n)}(x)\cdot dx - \lim_{n\to\infty}\int_0^d 1-F_{\mathbf{A}}^{(n)}(x)\cdot dx = 1-1=0. \qquad (26)
$$

The first integral on the the right hand side (RHS) converges to one by the following chain of identities:

$$\begin{aligned}\lim_{n\to\infty}\int_0^\infty 1-F_{\mathbf{A}}^{(n)}(x)\cdot dx &= \lim_{n\to\infty}\int_0^\infty 1-\frac{1}{n}\sum_{i=1}^{n}1\{\lambda_i<x\}\cdot dx\\ &= \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\int_0^\infty 1-1\{\lambda_i<x\}\cdot dx\\ &= \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\lambda_i\{\mathbf{A}\} = \lim_{n\to\infty}\frac{1}{n}\mathrm{Tr}[\mathbf{A}] = 1.\end{aligned}$$

The second term on the RHS of (26) is identified to converge to one by taking the limit inside the integral. Again, this can be done, since we deal with a Riemann integral over a uniformly convergent function. ■

Lemma 3. *Let* $\mathbf{A},\mathbf{B}\in\mathbb{C}^{n\times n}$ *be positive-semidefinite random matrices with LSMs* $f_{\mathbf{A}}(x)=\delta(x-1)$ *and* $f_{\mathbf{B}}(x)=\psi(x)$*, respectively. Then, the LSM of* $\mathbf{AB}$ *is given by* $f_{\mathbf{AB}}(x)=\psi(x)$*.*

Proof. We separate the eigenvalues $\mu_i \triangleq \lambda_i\{\mathbf{A}-\mathbf{I}_n\}$, $i\in\{1,\dots,n\}$, into two sets $\mathcal{L}_1$ and $\mathcal{L}_2$. For a fixed $\varepsilon>0$ the eigenvalues in the first set fulfill $|\mu_i|\le\varepsilon$. The second set contains the eigenvalues which fulfill $|\mu_i|>\varepsilon$. Firstly, we show that the eigenvalues in $\mathcal{L}_2$ do not have any impact on the LSM of $\mathbf{AB}$. Since $\mathbf{A}-\mathbf{I}_n$ is Hermitian, we can write with $\tilde{\mathbf{A}}\triangleq\mathbf{I}_n+\sum_{i:\mu_i\in\mathcal{L}_1}\mu_i\mathbf{v}_i\mathbf{v}_i^{\mathrm{H}}$

$$\mathbf{AB}=\tilde{\mathbf{A}}\mathbf{B}+\sum_{i:\mu_i\in\mathcal{L}_2}\mu_i\mathbf{v}_i\mathbf{v}_i^{\mathrm{H}}\mathbf{B},$$

where $\mathbf{v}_i$ denotes the eigenvector corresponding to μ_i. The EED of $\mathbf{A}$ a.s. converges to $\sigma(x-1)$. Therefore, the number of eigenvalues in $\mathcal{L}_2$ grows less than linearly in n. Since the $\mathbf{v}_i\mathbf{v}_i^{\mathrm{H}}$'s are unit rank matrices, we conclude that the fraction of differing eigenvalues of $\mathbf{AB}$ and $\tilde{\mathbf{A}}\mathbf{B}$ goes to zero as $n\to\infty$. Thus, they also share the same LSM.

Secondly, we show that the LSM of $\tilde{\mathbf{A}}^{-1}+\rho\mathbf{B}$ is given by

$$f_{\tilde{\mathbf{A}}^{-1}+\rho\mathbf{B}}(x)=\rho^{-1}f_{\mathbf{B}}(\rho^{-1}x-1). \tag{27}$$

Note that the eigenvalues of $\tilde{\mathbf{A}}^{-1}$ are the inverse eigenvalues of $\tilde{\mathbf{A}}$. Therefore, we can write $\tilde{\mathbf{A}}^{-1}=\mathbf{I}_n+\mathbf{\Delta}$, where a.s. for any $\delta>0$ there exists an n_0 such that for all $n\ge n_0$

$$\max_{i\in\{1,\dots,n\}}\lambda_i\{\mathbf{\Delta}\}<\delta. \tag{28}$$

Let's denote the normalized eigenvectors of $\tilde{\mathbf{A}}^{-1}+\rho\mathbf{B}$ corresponding to $\lambda_i\{\tilde{\mathbf{A}}^{-1}+\rho\mathbf{B}\}$ by $\mathbf{u}_i$. By the definition of an eigenvector, we can write

$$\left(\mathbf{I}_n+\boldsymbol{\Delta}+\rho\mathbf{B}-\lambda_i\{\tilde{\mathbf{A}}^{-1}+\rho\mathbf{B}\}\cdot\mathbf{I}_n\right)\cdot\mathbf{u}_i=\mathbf{0}$$

for $i\in\{1,\ldots,n\}$. Taking $\boldsymbol{\Delta}\mathbf{u}_i$ to the RHS and taking the Eukledian norm yields

$$\|\rho\mathbf{B}\mathbf{u}_i+(1-\lambda_i\{\tilde{\mathbf{A}}^{-1}+\rho\mathbf{B}\})\mathbf{u}_i\|=\|\boldsymbol{\Delta}\mathbf{u}_i\|. \tag{29}$$

By (28) and $\|\mathbf{u}_i\|=1$, we conclude that for for all $n\geq n_0$ a.s. also

$$\|\boldsymbol{\Delta}\mathbf{u}_i\|<\max_{i\in\{1,\ldots,n\}}\lambda_i\{\boldsymbol{\Delta}\}<\delta. \tag{30}$$

Thus, for all i the RHS of (29) goes to zero as $n\to\infty$. Accordingly, we conclude for the LHS that $\mathbf{u}_i\to\mathbf{w}_i$ and $\lambda_i\{\tilde{\mathbf{A}}^{-1}+\rho\mathbf{B}\}\to 1+\rho\nu_i$, where ν_i and $\mathbf{w}_i$ are the ith eigenvalue and eigenvector of $\mathbf{B}$. The respective variable transformation in the LSM of $\mathbf{B}$ yields (27).

We complete the proof by showing that the Shannon transforms [14], [15] of $f_{\tilde{\mathbf{A}}\mathbf{B}}(\cdot)$ and $f_{\mathbf{B}}(\cdot)$ coincide. Note that the Shannon transform contains the full information about the corresponding distribution. Consider the quantity $\xi\triangleq n^{-1}\log\det\left(\tilde{\mathbf{A}}^{-1}+\rho\mathbf{B}\right)-n^{-1}\log\det\left(\tilde{\mathbf{A}}^{-1}\right)$. As $n\to\infty$ this quantity converges to the Shannon transform of $f_{\mathbf{B}}(\cdot)$, $\Upsilon_{\mathbf{B}}(\rho)$, a.s.:

$$\begin{aligned}\lim_{n\to\infty}\xi&=\int_0^\infty\log x\cdot f_{\tilde{\mathbf{A}}^{-1}+\rho\mathbf{B}}(x)\cdot dx-\int_0^\infty\log x\cdot f_{\tilde{\mathbf{A}}^{-1}}(x)\cdot dx\\&=\int_0^\infty\log(1+\rho x)\cdot f_{\mathbf{B}}(x)\cdot dx-\int_0^\infty\log x\cdot\delta(x-1)\cdot dx\\&=\int_0^\infty\log(1+\rho x)f_{\mathbf{B}}(x)\cdot dx\triangleq\Upsilon_{\mathbf{B}}(\rho)\end{aligned}$$

Rewriting $\xi=n^{-1}\log\det\left(\mathbf{I}_n+\rho\tilde{\mathbf{A}}\mathbf{B}\right)$, we see that ξ also converges to the Shannon transform of $f_{\tilde{\mathbf{A}}\mathbf{B}}(\cdot)$, $\Upsilon_{\tilde{\mathbf{A}}\mathbf{B}}(\rho)$, a.s.:

$$\lim_{n\to\infty}\xi=\int_0^\infty\log xf_{\mathbf{I}_n+\rho\tilde{\mathbf{A}}\mathbf{B}}(x)\cdot dx=\int_0^\infty\log(1+\rho x)f_{\tilde{\mathbf{A}}\mathbf{B}}(x)\cdot dx\triangleq\Upsilon_{\tilde{\mathbf{A}}\mathbf{B}}(\rho).$$

Accordingly, we conclude that $f_{\mathbf{A}\mathbf{B}}(x)=f_{\tilde{\mathbf{A}}\mathbf{B}}(x)=f_{\mathbf{B}}(x)=\psi(x)$. $\blacksquare$

REFERENCES

[1] S. Borade, L. Zheng, and R. Gallager, "Maximizing degrees of freedom in wireless networks," in *Proc. Allerton Conf. Commun., Contr., and Computing*, Monticello, IL, Oct. 2003.

[2] ——, "Amplify and forward in wireless relay networks: Rate, diversity and network size," *IEEE Trans. Inform. Theory, Special Issue on Relaying and Cooperation in Communication Networks*, Oct. 2007, to appear.

[3] S. Yeh and O. Lévêque, "Asymptotic capacity of multi-level amplify-and-forward relay networks," in *Proc. IEEE Int. Symposium on Inf. Theory*, Nice, France, June 2007.

[4] J. Silverstein, "Strong covergence of the empirical distribution of eigenvalues of large dimensional random matrices," *Journal of Multivariate Anal.*, vol. 55, pp. 331–339, Nov. 1995.

[5] R. R. Müller, "On the asymptotic eigenvalue distribution of concatenated vector-valued fading channels," *IEEE Trans. Inform. Theory*, vol. 48, no. 7, pp. 2086–2091, July 2002.

[6] V. I. Morgenshtern and H. Bölcskei, "Random matrix analysis of large relay networks," in *Proc. Allerton Conf. Commun., Contr., and Computing*, Monticello, IL, Sept. 2006.

[7] ——, "Crystallization in large wireless networks," *IEEE Trans. Inform. Theory, Special Issue on Relaying and Cooperation in Communication Networks*, Oct. 2007, to appear.

[8] H. Bölcskei, N. U. Nabar, O. Oyman, and A. J. Paulraj, "Capacity scaling laws in MIMO relay networks," *IEEE Trans. Wireless Commun.*, vol. 5, no. 6, pp. 1433–1444, June 2006.

[9] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–595, Nov. 1999.

[10] V. A. Marcenko and L. A. Pastur, "Distributions of eigenvalues for some sets of random matrices," *Math. USSR-Sbornik*, vol. 1, pp. 457–483, 1967.

[11] Z. D. Bai, "Methodologies in spectral analysis of large dimensional random matrices," *Statistica Sinica*, vol. 9, pp. 611–677, 1999.

[12] D. Voiculescu, "Multiplication of certain non-commuting random variables," *Journal of Operator Theory*, vol. 18, pp. 223–235, 1987.

[13] R. R. Müller, "Random matrices, free probability and the replica method," in *Proc. of European Signal Processing Conference (Eusipco)*, Vienna, Austria, Sept. 2004.

[14] S. Verdú, "Large random matrices and wireless communications," in *MRSI Information Theory Workshop*, Berkeley, CA, Feb./Mar. 2002.

[15] A. M. Tulino and S. Verdú, *Random matrix theory and wireless communications*. Now Publishers Inc., 2004.

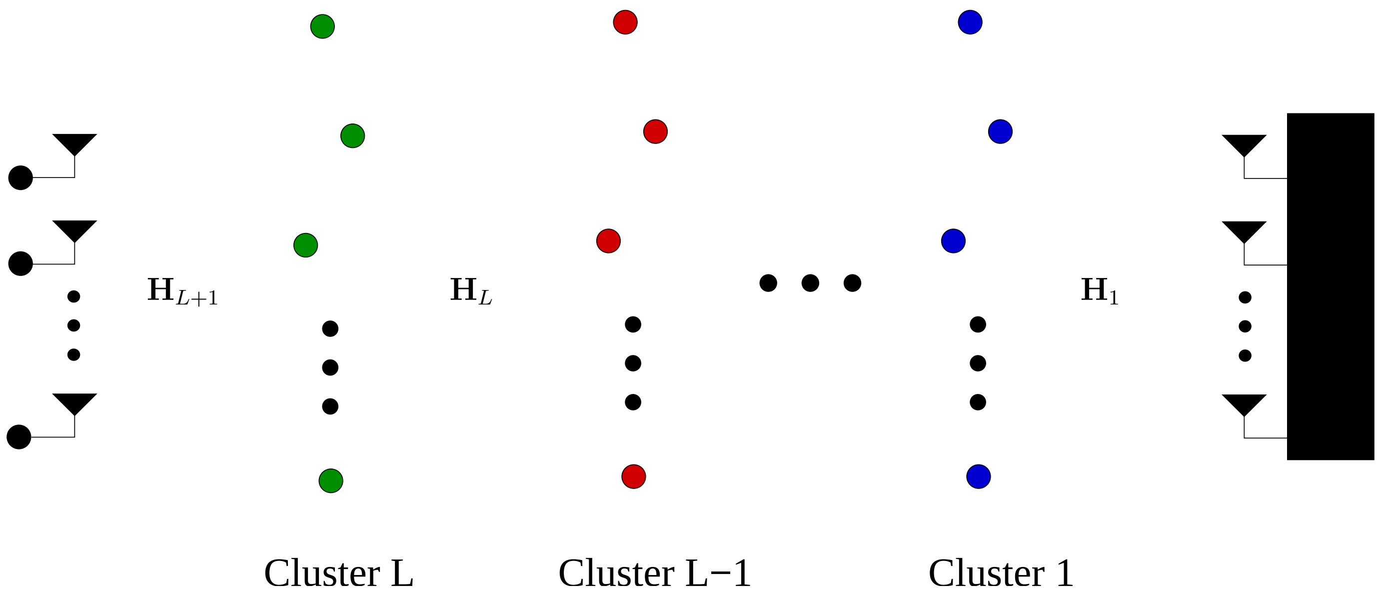

Fig. 1. n_{s} non-cooperating source antennas transmit to a destination terminal with n_{d} antennas via L clusters of k non-cooperating relay antennas.

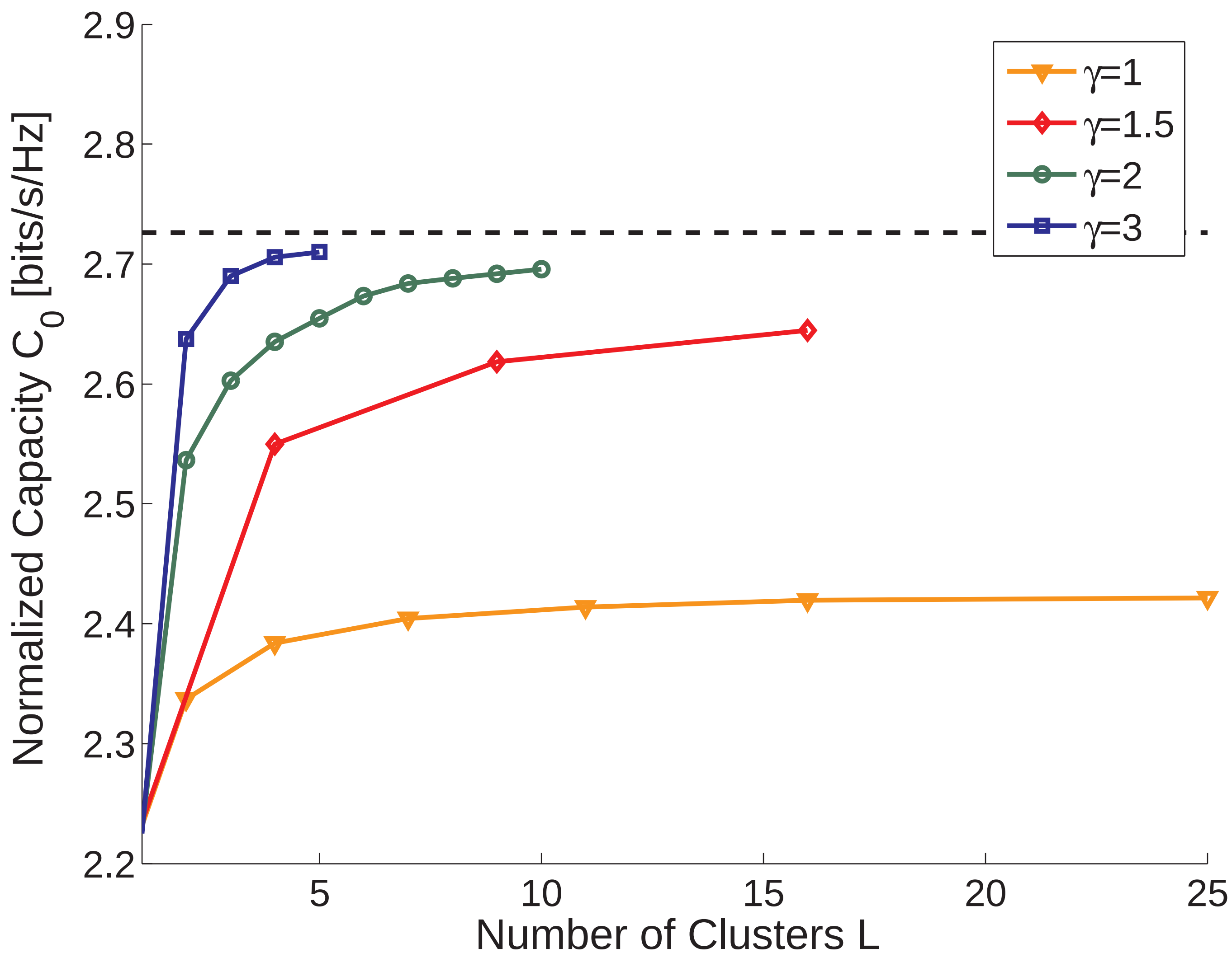

Fig. 2. Normalized capacity C_0 versus the number of relay clusters L as obtained through Monte Carlo simulations. The number of source and destination antennas is $n = 10$. The SNR at the destination is 10 dB. The number of relays per cluster k evolves according to $k = n \cdot L^{\gamma}$. The dashed curve shows the normalized point-to-point MIMO capacity as a reference.

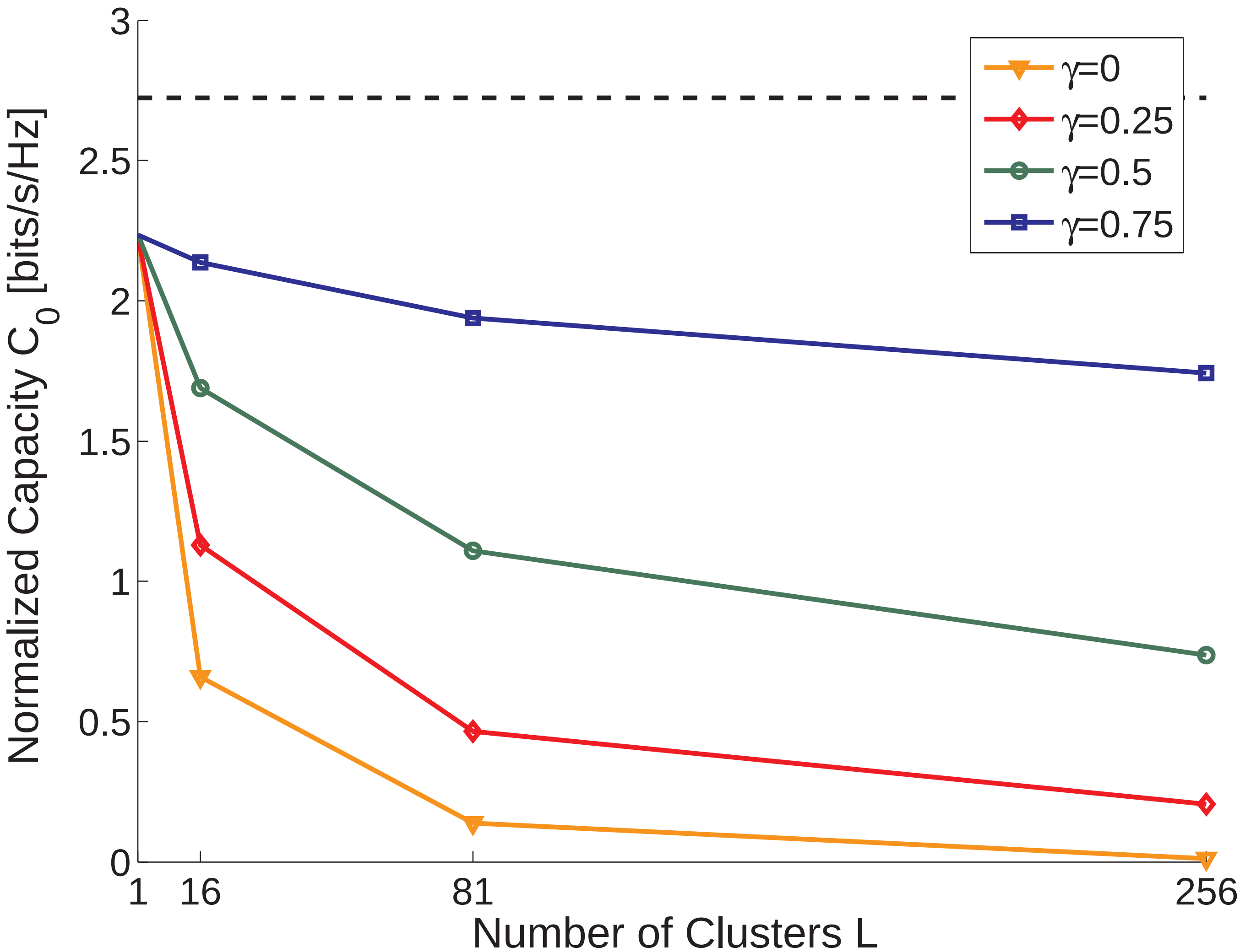

Fig. 3. Normalized capacity C_0 versus the number of relay clusters L as obtained through Monte Carlo simulations. The number of source and destination antennas is $n = 10$. The SNR at the destination is 10 dB. The number of relays per cluster k evolves according to $k = n \cdot L^{\gamma}$. The dashed curve shows the normalized point-to-point MIMO capacity as a reference.